MoonRobots Gate

原创 MoonBit 开源库与开发工具 | robots.txt 可解释策略审计与抓取规划

参赛者	石子硕	联系方式	已通过官方报名问卷提交
项目类型	原创 MoonBit 开源库与工具	公开仓库	https://github.com/2111950632/moonrobots-gate
Mooncakes	2111950632/robots-gate	许可证	Apache-2.0

核心库 4,310 行	MoonBit 总计 5,846 行	107 项四目标测试	12 个核心模块

项目简介

MoonRobots Gate 将 robots.txt 文本转换为可检查的数据模型，为爬虫、搜索索引器、链接检查器、文档工具和 AI 智能体提供确定的访问判断、风险诊断、站点地图校验与抓取规划能力。

一、最终交付基础

项目已完成 MoonBit 0.2.2 正式版本，包含防御式策略解析、用户代理选择、最长规则匹配、Allow 同长优先、通配符与末尾锚点、百分号编码归一化、决策轨迹、语义检查、策略渲染、访问矩阵、策略差异、站点地图校验和预算抓取规划。现有九组 CLI 命令、四个示例和 107 项测试；格式、类型、构建、四目标测试和发布包检查均已通过。

二、本次活动完成内容与范围调整

项目从解析器原型扩展为模块化 Robots 协议策略工程工具，完成稳定诊断码、语义审计、Markdown / JSON / CSV 输出、策略差异、站点地图冲突检查、预算抓取规划、九组 CLI、四个示例和四目标 CI。实现阶段将网络获取调整为后续独立包，使核心保持无网络副作用、确定性和跨目标可移植性。

三、技术路线

采用“纯策略核心 + 调用方传输层”的分层结构：文本先经过限额解析和源位置诊断，再进入用户代理选择与路径匹配；上层复用统一决策结果完成语义审计、版本差异、站点地图检查和抓取规划。HTTP、重定向、缓存、DNS 和持久化由调用方或未来独立传输包负责，核心不包含页面解析、JavaScript 渲染和完整网络爬虫。

四、已完成功能、测试与文档

已完成可解释策略判断、语义审计、结构化报告、完整 CLI、站点地图策略检查和预算抓取规划。107 项测试覆盖协议语义、异常输入、资源上限、渲染、差异、站点地图和规划，并在 wasm、wasm-gc、JavaScript、native 四个目标通过。已提供 README、架构说明、生态查重、开发复盘、更新日志、提交清单和可运行示例。

五、原创性与许可证说明

本项目为原创 MoonBit 实现，不是其他开源库的直接移植，也未复制第三方项目源码。协议行为参考 RFC 9309、RFC 3986 和公开的 Robots 解析文档，仅用于理解公开标准。项目采用 OSI 认可的 Apache-2.0 许可证，核心运行时依赖仅为 moonbitlang/core；新增依赖将在引入前核查许可证并记录来源。

报名及材料提交以本次 8 月黑客松飞书问卷和后续官方通知为准。